

STUART E. SCHIFFER
Acting Assistant Attorney General
SANDRA M. SCHRAIBMAN
CAROL FEDERIGHI
AMY M. ALLEN
Attorneys, U.S. Department of Justice
Civil Division, Federal Programs Branch
P.O. Box 883, 901 E Street, N.W.
Washington, D.C. 20044
Telephone: (202) 514-4523
Facsimile: (202) 616-8202
Attorneys for Defendants

FILED
CLERK, U.S. DISTRICT COURT

11 JUL 01 PM 3:54

DISTRICT OF UTAH

BY: [Signature]
DEPUTY CLERK

**IN THE UNITED STATES DISTRICT COURT
DISTRICT OF UTAH, CENTRAL DIVISION**

STATE OF UTAH, et al.,

Plaintiffs,

v.

**DONALD L. EVANS, United States Secretary
of Commerce, et al.**

Defendants.

Civil No. 2:01-CV-000292G

**DECLARATION OF
HOWARD HOGAN**

I, Howard Hogan, declare as follows:

1. I am the Chief of the Decennial Statistical Studies Division ("DSSD") of the Bureau of the Census of the United States Department of Commerce. My primary task for the past three and a half years has been to lead the group of mathematical statisticians and programmers who are responsible for the statistical design of the United States Census. This responsibility includes overseeing the use of statistical methods in the decennial census, including imputation, statistical quality control design, coverage improvement activities, software quality

29

control, and content sampling, as well as the count sampling activities not related to apportionment. I am submitting this declaration in support of Defendants' Motion to Dismiss or, in the Alternative, Cross-Motion for Summary Judgment in the above-captioned action. This declaration is based on personal knowledge and on information provided to me in the course of my official duties.

2. A copy of my curriculum vitae ("C.V.") is attached to this declaration as Exhibit

A. As stated therein, I obtained my Bachelor's Degree in Economics/Mathematics from Pomona College in 1971, a Certificate in Economics from Stockholm University in 1972, a Master's Degree in Public Affairs from the Woodrow Wilson School at Princeton University in 1974, a Master's Degree in Economics from Princeton University in 1975, and a Ph.D. in Demography from Princeton University in 1976, issued by the Economics Department of that University.

3. I am a member and former Chair of the Nominating Committee of the International Association of Survey Statisticians, a member and former Representative at Large of the Washington Statistical Society, and a member of the Population Association of America. I am also a member and Program Chair-Elect of the Section on Survey Research Methods of the American Statistical Association ("ASA"). I was recently elected a Fellow with the ASA, a designation that recognizes an individual's outstanding service to and leadership in the field of statistical science. Each year, ASA's Committee on Fellows, a peer group, can only elect one-third of one percent of the total membership as Fellows.

4. I have been employed as a Professorial Lecturer in the Department of Statistics at George Washington University since August 1997, where I have taught applied survey sampling

and elementary statistics. From 1984 to 1996, I was an Instructor in the U.S. Department of Agriculture Graduate School Evening Program, where I developed the curriculum and taught a course on exploratory data analysis. I have also taught versions of this course at Statistics Sweden (1989) and as a Washington Statistical Society Short course (1996). I have additionally worked as a Visiting Scholar at the Department of Biostatistics at the University of North Carolina, as a Visiting Research Associate in the Department of Statistics at Macquarie University, and as the Head of the Demographic Unit at the University of Dar es Salaam. I have also published in the field of statistics. A list of my publications is included on the attached C.V.

5. I have been employed by the Census Bureau since 1979 in positions of increasing responsibility. From 1979 to 1983, I was employed in the Statistical Research Division, where I was responsible for researching the application of statistical and demographic techniques to census coverage measurement, attaining the position of Principal Researcher. In 1983, I was promoted to Chief of the Undercount Research Staff in the Statistical Research Division, where I led the Census Bureau's effort to research, develop, and implement methods to correct the 1990 Census for undercount. From 1993 to 1997, I was an Assistant Division Chief for Research and Methodology in the Services Division, where I was responsible for developing and conducting a program of research designed to make the broadest possible application of statistical principles and theories to the Census Bureau's surveys of retail trade, wholesale trade, service industries, and transportation. Besides the United States census, I have worked on the censuses of Morocco and Tanzania and additionally have consulted at the national statistics agencies of Mauritius, India, Canada, Sweden, and the United Kingdom.

Summary and Conclusions

6. The purpose of this declaration is to describe the use of imputation in Census 2000 and previous censuses. In preparing this declaration, I have reviewed the Complaint in this action, the memorandum in support of plaintiffs' motion for summary judgment, and the two declarations filed by the plaintiffs, the Declaration of Donald B. Rubin, Ph.D., and the Declaration of Lara J. Wolfson, Ph.D. I have also reviewed two affidavits, those of Peter Bounpane (former Assistant Director for Demographic Censuses at the Census Bureau) and Barbara Bailar (former Associate Director for Statistical Standards and Methodology), that were filed by the Census Bureau in a prior lawsuit challenging the use of count imputation, Orr v. Baldrige, No. IP 81-604-C, slip op. (S.D. Ind. July 29, 1985), along with Defendants' Answers to Plaintiffs' First Set of Interrogatories and an affidavit filed by Dr. Rubin in that lawsuit. The Bounpane and Bailar Affidavits are included in the Administrative Record ("A.R.") at C00566-91, the Orr opinion is at A.R. C00654-65, and the interrogatory responses are at A.R. C01161-75. Though I may differ in some of the details and choice of language, I am in basic agreement with the statements made by Mr. Bounpane and Dr. Bailar in their affidavits and with the interrogatory responses.

7. My conclusions are as follows:

- Imputation is a longstanding, valid, and widely-recognized statistical method used to account for missing, contradictory, or improperly processed data. Count imputation has been used as part of the Census Bureau's traditional enumeration process in every decennial census since (and including) 1960.

- Imputation and sampling are two separate, distinct statistical methods.
Imputation is not sampling.
- The Census Bureau used count imputation in Census 2000 because research and experience have confirmed that this technique improves census accuracy. While the Bureau is still evaluating the imputation results along with many other aspects of Census 2000, the Census Bureau has no reason to question its longstanding conclusion that count imputation improves accuracy.

Imputation Generally

8. Imputation is the process of assigning plausible values for missing data in censuses or surveys.¹ Bailar Aff. ¶ 12 (A.R. C00570); Orr, slip op., at 5 (A.R. C00658); D. Harner, Imputation in the 1980 Census (Nov. 4, 1982) (A.R. C00616-27). The Bureau used imputation in the 2000 decennial census as it has in several prior censuses to address the problem of missing, incomplete, and contradictory data, an inevitable aspect of censuses and surveys. The goal in Census 2000, as in all prior U.S. censuses, was to secure a complete and accurate count of all residents. Kenneth Prewitt, Director, U.S. Bureau of the Census, Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000, at 1 (June 2000) (A.R. C01904) (hereafter "Feasibility Statement"). It is, however,

¹ The Census Bureau usually uses the term "survey" or "sample survey" to refer to a data collection effort based on a sampling methodology, where the sample taken is generally a relatively small fraction of the total population being studied (e.g., 0.05%, 0.3%, 10%, 25%).

impossible to achieve this goal perfectly in an endeavor as massive and complex as a decennial census. In Census 2000, the Census Bureau processed data for over 120 million households, including over 147 million paper questionnaires and 1.5 billion pages of printed material. The workload during peak operations was about 3.3 million forms per day. Final Data Receipt Summary (Sept. 15, 2000) (A.R. C01578). Given the size of this undertaking, it would be unreasonable to expect the amount of missing or discrepant data to be zero – some amount of missing data is simply unavoidable. D. Harner et al., Imputation in the 1980 Census 1 (Nov. 4, 1982) (A.R. C00616).

9. As used in this Declaration, the phrase "missing data" refers to a wide variety of situations. The categories of missing data and the processes that lead to them have changed over the decades as the census has become increasingly computerized and centralized. In the 2000 Census, the various situations that resulted in missing data included incomplete or unavailable responses from housing units with previously confirmed addresses, conflicting data from the same housing unit, and failures in the data-capture process. The various types of missing data included characteristic data (information about an enumerated person, such as sex, race, age), population count data (information about the number of occupants in an identified housing unit), and housing unit status data (whether the unit is vacant, occupied, or nonexistent). As I will explain further below in paragraphs 15, 58-60, and 68-69, in Census 2000, the missing population-count and housing-unit-status data all arose from housing units whose addresses were recorded on the Bureau's Decennial Master Address File ("DMAF").

10. In the decennial census, the basic alternatives for handling missing data for an

identified address are either to tabulate the address as "data not reported," that is, to assign no value to the empty data fields, or to assign plausible values for the missing data.² See J. Beresford, The Need for a Study of Persons Created by Computer Edit in the 1960 Census 1 (Jan. 2, 1963) (A.R. C01188) (hereafter "Beresford"); M. Lear, Occupied Units Misclassified as Vacant, and Erroneous Deletions (Feb. 19, 1968) (A.R. C01197). In both cases, imputation takes place, either by default or explicitly. Bailar Aff. ¶¶ 3-4 (A.R. C00567); D. Harner et al., Imputation in the 1980 Census 1 (Nov. 4, 1982) (A.R. C00616); C. Jones, 1980 Census Imputation of Population Characteristics 2 (Feb. 5, 1988) (A.R. C01258). The first alternative is the equivalent of imputing a value of zero, i.e., of deciding that all returns with questionable or incomplete data represent vacant or nonexistent housing units. This conclusion is demonstrably untrue, as studies have shown that a significant proportion of returns with questionable or incomplete data or unresolved status are actually valid, occupied housing units. Id.; see also Census Undercount Adjustment: Basis for Decision, 45 Fed. Reg. 69,366, 69,373 (Oct. 20, 1980) (A.R. C01220); B. Denton, Results of National Vacancy Check (Apr. 30, 1971) (A.R. C00472-80); J. Curry, Background Information on Indiana Suit (July 28, 1981) (A.R. C00555-57); D. Whitford, DSSD Census 2000 Procedures and Operations Memorandum Series B-3, Quality of Census 2000 Processes (Prototype) 45-46 (Sept. 27, 2000) (A.R. C00928-29) (hereafter "2000 Quality Prototype (B-3)"). Under the second alternative, a value other than zero is inserted in the empty field. Bailar Aff. ¶ 12 (A.R. C00570). Because incomplete information may distort census

² Other alternatives, not relevant to the present dispute, exist for addressing missing data in other, noncensus contexts.

results, the issue is not whether to impute, but what type of imputation will be the most accurate.

11. The Census Bureau considers imputation part of its traditional census-taking methodology. See Census Undercount Adjustment: Basis for Decision, 45 Fed. Reg. 69,366, 69,373 (Oct. 20, 1980) (A.R. C01220); Census 2000 Operational Plan Using Traditional Census-Taking Methods II-6, XI-4 (Jan. 1999) (A.R. C01752, C01818) (hereafter "Census 2000 Operational Plan (Jan. 1999)"); Updated Summary: Census 2000 Operational Plan 9 (Feb. 1999) (A.R. C01731) (hereafter "Census 2000 Update (Feb. 1999)"). The decennial census has never been a simple physical "headcount" of inhabitants. The conduct of the decennial census has routinely involved techniques designed to obtain and use reliable information from a variety of sources concerning the aggregate number of persons residing at particular locations. For example, from the beginning, the census-takers relied on information provided by a member of the household, rather than on their own direct observation. Act of Mar. 1, 1790, ch. 2, § 1, 1 Stat. 101; General Government Division, United States General Accounting Office, GAO/GCD-98-103, Decennial Census: Overview of Historical Census Issues 15-16, 24 (May 1998) (Ex. B hereto). And, when a member of the household could not be found, census-takers have historically relied on information from neighbors, landlords, postal workers, or other proxies. Id. at 24-25; Bounpane Aff. ¶ 12 (A.R. C00577); United States General Accounting Office, Testimony, GAO/T-GGD-91-8, Statement of L. Nye Stevens, Director, Government Business Operations Issues, General Government Division, Before the Subcomm. on Census and Population, House Comm. on Post Office and Civil Serv., 102d Cong., 1st Sess. 9 (Feb. 21, 1991) (A.R. C00676) (hereafter, "Nye Statement"). This procedure was first formally authorized

for the 1880 census. See Act of March 3, 1879, ch. 195, § 8, 20 Stat. 473, 475.

12. The Census Bureau uses two primary types of imputation. The first type, characteristic imputation, is not at issue in this case. Characteristic imputation supplies missing characteristic data for a housing unit's response, but does not concern the population count. For example, if a given housing unit did not provide ages for the individuals living in the housing unit, but supplied all other information, age would be imputed for the individuals in that housing unit.

13. The second type of imputation, which I have called "count imputation,"³ is imputation that assigns a population count to a housing unit. Count imputation procedures apply when the Census Bureau is unable to secure any information regarding a given address, or when the Census Bureau has limited or contradictory information about the address but not the definitive number of occupants.

14. Count imputation was used in three situations in Census 2000. See DSSD Census 2000 Procedure and Operations Memorandum Series #Q-34, Census 2000 Specifications for Imputing Housing Unit Status and Population Counts (Sept. 26, 2000) (hereafter "2000

³ The Census Bureau has not always used the term "count imputation" to denominate this type of imputation. In fact, it has used a variety of different terms that may partially account for the misunderstanding of the application of this methodology. See J. Clark, Definitions for 1990 Imputation Evaluation 2 (Nov. 5, 1992) (A.R. C00730). For example, some Census Bureau terms group other types of imputation with count imputation. The terms "substitution" and "whole-person imputation" include "count imputation," as well as imputation of a full set of characteristics (age, sex, etc.) for a known, enumerated individual. See S. Love et al., Definitions of Substitution (May 9, 2001) (A.R. C01138-45) (substitution); Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy 25 (Mar. 1, 2001) (A.R. C01061) (whole-person imputation).

Imputation Specifications (Q-34)) (A.R. C00858-86):⁴

- Household Size Imputation – The Census Bureau imputed a population count for a housing unit when Bureau records indicated that the housing unit was occupied, but did not show the number of individuals residing in the unit.
- Occupancy Imputation – When Census Bureau records indicated that a housing unit existed but not whether it was occupied or vacant, the Bureau imputed occupancy status (occupied or vacant), and then, if the unit was imputed to be occupied, imputed a household size, as above.
- Status Imputation – When the Census Bureau's records had conflicting or insufficient information about whether an address represented a valid, non-duplicated housing unit, the Bureau first imputed for the status of the unit (occupied, vacant, delete), then, if occupied, the household size of the donor record is used.

15. With regard to the use of status imputation in 2000, the Bureau did not use this methodology on fictional or "phantom" addresses, as plaintiffs and Dr. Wolfson assert. Status

⁴ Again, the Census Bureau has not used consistent terminology in the past to denominate these three types of count imputation (household size imputation, occupancy imputation, and status imputation) and, in fact, has not often had a reason to define specific terms. See DSSD Census 2000 Procedures and Operations Memorandum Series B-1*, Data and Analysis to Inform the ESCAP Report 21-22 (Mar. 1, 2001) (A.R. C01089-90) (not using specific terminology); 2000 Imputation Specifications (Q-24) (A.R. C00858-86) (same). I have chosen to use these terms in the Declaration because I think they most clearly describe the processes at issue.

imputation was applied to addresses included in the Bureau's DMAF. Addresses in the DMAF originated from official sources such as the 1990 Census, the U.S. Postal Service, local governments, tribal governments, and Census Bureau enumerators. These addresses were validated and updated according to specific Bureau procedures. See The Census Bureau's Master Address File (MAF): Census 2000 Address List Basics (Mar. 1999) (A.R. C01490-504); DSSD Census 2000 Procedures and Operations Memorandum #D-6, Specification for Updating the Decennial Master Address File in April, 2000 (Mar. 27, 2000) (A.R. C01545-54). Because the Census Bureau had secured sufficient information to include these addresses in the DMAF, these are not "phantom" addresses.

16. The basic method the Census Bureau has used for count imputation since the 1960 census is the "hot-deck" methodology, in which the imputed information comes from the same census. Encyclopedia of the U.S. Census 196 (M. Anderson ed., 2000) (A.R. C00430) (hereafter Encyclopedia). "Cold-deck" imputation uses information from a prior census or some other outside source. Id. The hot-deck methodology has the advantage of using contemporaneous data, as well as of preserving a realistic distribution of data. Id.

17. The hot-deck method that was used is premised on local homogeneity, i.e., the fact that housing units in close geographic proximity tend to be similar in most respects, including size. See ¶ 31 infra. In the sequential form of hot-deck imputation used in the earlier censuses, housing unit data were stored sequentially in a file as they were processed. Encyclopedia at 196 (A.R. C00430). When occupancy or population data for a particular unit were incomplete, the computer referred to this sequential file, or hot deck, to obtain data from the most recently

processed housing unit. Id. In more recent censuses, this methodology has been improved to incorporate other factors believed to help identify the "donor" unit most similar to the unit subject to imputation, see DSSD 1990 REX Memorandum Series #BB-11, Summary of the 1990 Decennial Census Edit and Imputation Procedures for the 100% Population and Housing Items 3 (Oct. 5, 1994) (A.R. C00778) (hereafter "1990 Imputation Specifics (BB-11)"), and, in the 2000 Census, allowed the computer to search all housing units within a tract, not just the most recently processed one. See 2000 Imputation Specifications (Q-34) (A.R. C00858-86).

Imputation is not sampling

18. Statistics is a broad science. Many methodologies are included under the statistical umbrella, including sampling, imputation, weighting, etc. Imputation is a statistical methodology, but it is not sampling.

19. A review of how sampling and imputation relate, through an examination of the entirety of data collection and processing, demonstrates that sampling and imputation are not the same. When one wants to produce data describing a target population or subgroup, a set of activities is conducted that can be split into two stages: (1) collecting the data, and (2) processing the data. Both collection and processing can similarly be broken into smaller parts. Sampling may or may not be used in the first stage (data collection), while imputation is only one of several activities in the second stage (processing). The usages and purposes of imputation are therefore distinct from those of sampling.

20. When the Census Bureau conducts a census or a survey, it usually proceeds through the following two-step process:

21. Data Collection: First, the Bureau decides whom it will attempt to contact, i.e., everyone in the target group or a sample. If the Bureau is taking a sample survey, it then determines the sample design, that is, how to select the people or units to be contacted. Other aspects determined at this stage include the following: the time period for collecting the data, the questionnaire to be used, how the survey or census takers will deal (in the field) with nonrespondents, etc. Generally, the Bureau prepares a list of the units eligible to be contacted (the "frame") and, then, if a sample is being taken, the units to be included in the sample survey are selected from this frame according to the sample design. In a census, all units in the frame are included. The data are then collected through interviews, either in person, over the telephone, via mail, or through other modes. Finally, for most censuses and sample surveys, follow-up efforts must be made to recontact people who do not respond initially. These efforts might entail visiting a person who did not reply by mail or making follow-up telephone calls.

22. Data Processing: The second stage is data processing. Whether one is dealing with a census or a survey, first, the data are collected and input into a centralized database. Then, the data are "cleaned" through a variety of methods. One such method, imputation, is used to assign a value to an empty field on an otherwise completed questionnaire or to complete a blank questionnaire, to form a complete record for a given respondent. The imputation method selected may rely on prior experience or on data already collected to replace the missing data. Editing also occurs at this stage for responses that fail checks to ensure that the data on the entire questionnaire are consistent. (For example, the response could indicate that a person is two years old and also the spouse of the householder.) The inconsistent or incomplete data is replaced with

an imputed value that is consistent with other responses on the questionnaire. The data are then tabulated to determine totals, means, proportions, ratios, or other statistics. For sample surveys, weights that incorporate the probability of selection and possibly other adjustments are usually used in these computations.

23. Thus, the term sampling refers to one strategy of data collection and can be said to encompass some or all of the activities in the data collection stage of, e.g., a sample survey. By contrast, imputation (count imputation or other types) is a statistical procedure applied only in the second stage, the data processing stage. Thus, as Dr. Bailar correctly stated, sampling and imputation "are two completely different procedures, based upon totally distinct principles and serving equally distinct purposes." Bailar Aff. ¶ 12 (A.R. C00570). I also agree with Dr. Bailar's statement that sampling and imputation are not competitive, nor is one methodology superior to the other. Id.

24. The Census Bureau has always taken the position that imputation and sampling are different. The affidavit of Barbara Bailar, filed in the Orr lawsuit in 1982, demonstrates that the Census Bureau's longstanding position is that imputation and sampling are different.

25. Both Dr. Wolfson and Dr. Rubin rely on definitions of sampling to answer the question of whether imputation is sampling. Neither definition proves their assertion that imputation is sampling. Dr. Wolfson (¶ 45) defines sampling as "the process of selecting a number of subjects [units] from all the subjects [units] in a particular group or universe" (quoting from John M. Last, Ed., A Dictionary of Epidemiology, 3rd Ed., Oxford University Press, 1995 at 151, quotation marks deleted). She adds that "[i]n contemporaneous usage . . . 'statistical

sampling' is taken not only to be the process of selecting the subjects, but the subsequent use of the sample to draw inference to the population from which it was drawn."

26. First, Dr. Wolfson's definition has no application to the 2000 Census, where the Census Bureau attempted in numerous ways to contact each and every resident and made no "selection" of a subgroup in designing the Census. Second, although the definition more or less correctly describes sampling at a very high level of generality, it is not sufficiently precise, as Dr. Wolfson's later application of the definition shows. Her definition does not make clear that, in sampling, the process of selecting a sample is a deliberate and purposeful activity occurring during the design phase of a survey. Without this understanding, Dr. Wolfson's definition is broad enough to cover situations that have nothing to do with sampling. And with this understanding, Dr. Wolfson's definition does not encompass imputation. As explained above, imputation is not a mechanism for selecting units during the design phase of a census or sample survey, but rather is a means of dealing with missing data in the data processing stage.

27. Dr. Rubin (§ 13) defines sampling as follows: "Sampling refers to the process of obtaining data from a subset of a population (the subset is usually called the 'sample') from which estimates are made about characteristics of the entire population." This definition suffers from the same flaws as Dr. Wolfson's in that Dr. Rubin's definition does not incorporate the process of *deliberately selecting* a subset of the population *during the design phase of a survey*. Indeed, Dr. Rubin apparently recognizes this problem with his definition, because later in his declaration (§§ 16, 17) he writes that the "way in which the sample is chosen is critical" and that "[i]n order to use a sample for valid scientific estimation, specific sampling design requirements must be adhered

to." Without the notion of deliberate or purposeful selection according to a sampling design, however, Dr. Rubin's definition is as overly broad as Dr. Wolfson's. Under their definitions, all censuses in this or probably any other country have been "samples," as none to my knowledge has ever included each and every person in the population.⁵

28. Dr. Wolfson also claims (§ 47) that imputation and the methodology struck down in House are both statistical sampling because both "attempt to estimate the number of persons in households that were not actually enumerated by traditional census methods on the basis of statistical methods and assumptions that allow inferences to be drawn about an unobserved segment of the population based on information in an observed segment of the population." Basically, she is defining sampling as a process that allows inferences to be drawn about an unobserved segment of the population based on information in an observed segment of the population. As explained in the preceding paragraphs, this definition is entirely too broad.

29. There are many different kinds of sampling, including probability sampling and quota sampling. Dr. Rubin concedes (§ 22) that imputation is not probability sampling. See also Rubin Aff. (Orr) § 7 (Exhibit G to Defendants' Motion to Dismiss or, in the Alternative, Their Cross-Motion for Summary Judgment). Neither is imputation as used in the 2000 Census quota

⁵ A definition from a standard textbook is both more traditional and more useful than those of Dr. Rubin and Dr. Wolfson: "A sampling method is a method of selecting a fraction of the population in a way that the selected sample represents the population." Pandurang V. Sukhatme, Sampling Theory of Surveys With Applications 9 (1954). This definition emphasizes the fact that sampling refers to the deliberate selection of less than the whole population. Sukhatme's book was published around the time that 13 U.S.C. § 195 was first enacted and thus provides a useful context for the use of the word "sampling" in that statute.

sampling, as Dr. Rubin asserts (§ 17). In quota sampling, interviewers are sent out to get a prescribed number of interviews, usually divided by category (such as race, sex, etc.), but the particular choice of people, households, or housing units to interview is left to the discretion of the interviewer. For example, an interviewer might be sent out in a city to talk to a given number of young men, so many young women, so many African Americans, and so many Hispanics. These are the quotas. The interviewer can find and interview whichever households or people she or he wants as long as these quotas are met. The procedures used by the Census Bureau in its decennial censuses meet none of these criteria of quota sampling. In the Census, census takers are sent (repeatedly if necessary) to assigned housing units, with the task of getting interviews from the people living there. The enumerators are not allowed to choose the housing units that they visit or to interview housing units other than the ones to which they were assigned, and they are not given any quotas regarding the age, sex, race, or other characteristics of people to be interviewed.

30. As stated in the Census Bureau's interrogatory responses in Orr, when used in the context of statistical, demographic, and population survey applications, the term "sampling" generally refers to "probability sampling." A.R. C01167. The definition of sampling given by Dr. Bailar in her affidavit in Orr is an accurate description of probability sampling. Bailar Aff. § 12 (A.R. C00570) (Sampling is "the selection of a subset of units from a larger population in such a way that each unit of the population has a known chance of selection. Sampling is used where a scientifically selected set of units can be used to represent the entire population from which they are drawn and inferences to the entire population can be based on sample results."). As I

indicated above, there are other forms of sampling. "Sampling" in connection with the census (i.e., the asking of certain questions to only a fraction of the population) began in 1940 with probability sampling. E. Goldfield, National Research Council, Innovations in the Decennial Census of Population and Housing: 1940-1990, at 19 (Aug. 1992) (A.R. C01362) (hereafter "Innovations"). As even Dr. Rubin concedes in both of his declarations, the imputation described in this declaration and in the plan for Census 2000 is not probability sampling because it is unrelated to the selection of a subset of units from a larger population in such a way that each unit of the population has a known chance of selection.

Imputation Furthers Accuracy

31. Imputation is recognized in the statistical community as a procedure that can improve the accuracy and reliability of censuses and surveys, and has been amply discussed and documented in the professional literature. As Dr. Bailar stated in her affidavit in Orr, count imputation "has considerable support in both scientific research in other areas and in the Bureau's own expertise in conducting prior censuses and surveys." Bailar Aff. ¶¶ 18-23 (A.R. C00572-74), and studies cited therein; see ¶ 51 infra. The hot-deck method is recognized as a valid method for imputing in many circumstances, particularly in such large-scale censuses as the decennial census. Id.; Innis G. Sande, Hot-Deck Imputation Procedures, in 3 Incomplete Data in Sample Surveys 339 (W.G. Madow & I. Olkin eds., 1983) (Ex. C hereto); Frederick J. Scheuren, Discussion, in id. at 357 (Ex. D hereto); Barry L. Ford, An Overview of Hot-Deck Procedures, in 2 Incomplete Data in Sample Surveys 185 (W.G. Madow & I. Olkin eds., 1983) (Ex. E hereto). The homogeneity assumption that underlies the hot deck technique is also well supported in practice,

in the academic literature, and in logic. K. Wolters, Intraclass Correlation Study – Analysis of 1976 RAV Data (July 31, 1978); D. Harner, Results of Imputation Run Length Tabulation (Sept. 17, 1982) (A.R. C00601-06); D. Harner, Results of Imputation Run Length Tabulation for Special Study EDs (Sept. 20, 1982) (A.R. C00607-15); K. Thomas et al., Intraclass Correlations Using a Sample of 1980 Census Data (Jan. 31, 1984) (A.R. C00634-50).

32. Imputation makes the census more accurate because refusing to impute ignores some people that attempted to participate in the census. See ¶ 10 supra. Data regarding the returns from residents of housing units with conflicting or missing data would be excluded from the census, despite these individuals' attempts to be included. Even extensive attempts to obtain direct information from households will occasionally fail; in those instances imputation permits these individuals to be accounted for in the census.

33. Two types of accuracy are usually relevant to the issue of census accuracy. Numeric accuracy refers to how close the overall count of a particular geographic area or demographic group is to the "true" number of people who reside in that area or belong to that group. Distributive accuracy refers to how close the relative proportion or share of a geographic area or demographic group is to its true share relative to other areas or groups. A census operation that increases numeric accuracy moves the overall count for any particular area or demographic group closer to the true total. A census operation that increases distributive accuracy will improve the accuracy of the population share for a given area or demographic group compared to other areas or demographic groups – in other words, improve the accuracy of the estimated proportions or shares of the total population for the areas or groups. A perfect census

– one in which every resident is counted once and only once and is correctly located – would be both numerically and distributively accurate. Feasibility Statement, at 15-18 (A.R. C01917-20).

34. The goal in Census 2000 was to conduct a census that was both numerically and distributively accurate. This said, numeric accuracy drove the process for designing Census 2000 operations. When it designs any decennial census, the Census Bureau has available a very large number of possible operations. It assesses these operations against such criteria as cost, statutory deadlines, whether the staff necessary to implement these operations can be recruited and adequately trained, and how well the operations fit with other operations under consideration. In this extensive process of evaluating individual operations and then assembling them in the final design, there is one paramount criterion: what census design has the highest probability of correctly enumerating the population? That is, can an operation considered separately, and when combined with other operations, be expected to help the Census Bureau correctly count as many people as possible, given funding, timing, and other constraints? However, because it is difficult and perhaps impossible to know *a priori* the effects of a particular census operation on distributive accuracy, assessing an operation's effect on distributive accuracy can rarely be part of the planning process. Feasibility Statement, at 15-18 (A.R. C01917-20).

35. In principle, any given census operation designed to increase numeric accuracy can increase distributive accuracy, leave it the same, or make it worse. In designing Census 2000, the Census Bureau did not reject operations that would improve numeric accuracy (and meet other criteria for inclusion) even if these operations might affect distributive accuracy negatively, or indeterminately. For example, the Census Bureau developed for Census 2000 an extensive

partnership program to assist local jurisdictions and community organizations in promoting participation in the Census. But increasing the counts for these participating localities might not necessarily have translated into improvements in distributive accuracy. If one state promoted the Census more effectively than another state, the state with the better promotion program might have earned a higher share of the total national population than would otherwise be the case. And, although the Census Bureau largely targeted its coverage improvement programs in the areas that have been the most difficult to count, it did not reject census operations that might disproportionately have improved the count for groups that are already well counted. The Census Bureau viewed these increases in numeric accuracy, even for well counted groups, as important to the most basic goal of the Census – counting everyone.

36. Accordingly, in planning and designing Census 2000, the Census Bureau focused on achieving numeric accuracy. As explained above (§§ 10, 32), the use of count imputation furthered this goal by imputing a nonzero value for housing units, some of which demonstrably are occupied.

37. Dr. Wolfson contends (§§ 58-61) that the Census Bureau's imputation procedure is biased because it tends to overestimate the number of imputed persons. As I have stated, rather than increasing bias, the imputation process increases accuracy for the reasons set forth above. To impute zero, as Dr. Wolfson apparently would do, would actually result in *more* error, rather than less.

38. In addition, as with all census procedures, the Bureau is continually revising its procedures to reduce errors and other problems. For example, as discussed further below in

paragraph 59, the imputation procedures for Census 2000 required the donor household to be from an enumerator-completed form, which meant that in mailback areas donors were restricted to households requiring field follow-up. Also, if possible, the donor and donee were both from either a single unit structure or a multi-unit structure. 2000 Imputation Specifications (Q-34) (A.R. C00858-86). This change was implemented to reduce the kinds of biases that concerned Dr. Wolfson.

History of Imputation in the Decennial Census

39. The Census Bureau has used imputation to account for missing data since 1940, when it began characteristic imputation by imputing missing ages. R. Jenkins, Procedural History of the 1940 Census of Population and Housing 67 (1983) (A.R. C00344); Encyclopedia, at 197 (A.R. C00431); Innovations, at 45 (Aug. 1992) (A.R. C00719). The Census Bureau first used count imputation to add individuals to the census apportionment totals in the 1960 Census. 1960 Censuses of Population and Housing: Procedural History 81-84 (1966) (A.R. C00376-80) (hereafter "1960 Procedural History"); Encyclopedia, at 197 (A.R. C00431). Count imputation has been used in every census since 1960. See ¶¶ 41-60 *infra*. As demonstrated by many of the materials in the Administrative Record, the Census Bureau's use of imputation in the decennial censuses since 1940 has been made known to outside statisticians, Congress, and the courts, without any serious criticism being raised about the practice in over six decades, other than in the Orr case, which was resolved in the Census Bureau's favor.

40. The following is a more detailed description of how count imputation was used from 1960 to 1990:

1960 Census

41. The 1960 census was the first census to use computers to produce the apportionment count. See 1960 Procedural History, at 81-84 (A.R. C00376-80); Encyclopedia, at 197 (A.R. C00431). It was also the first census to use count imputation. Id. A portion of the individuals added through count imputation in the 1960 Census appear to have been necessary because of mechanical difficulties, specifically difficulties with FOSDIC (Film Optical Sensing Device for Input to Computers), the computer imaging device used to record questionnaires. 1960 Procedural History, at 83 (A.R. C00379). In general, the missing data were taken from the previously processed schedule, which usually represented a neighboring housing unit (the sequential hot-deck method). Id. The Bureau reported that "[p]ersons substituted" due to "noninterview" or "mechanical failure" were responsible for 0.5 percent of the total U.S. census count in 1960. Id. at 1-195 (Table B-1) (A.R. C00383). Because, as stated in note 4 supra, the Bureau's use of the term "substitution" has at various times included imputation of a full set of characteristics (age, sex, etc.) for a known, enumerated individual, in addition to count imputation, it is unclear whether the numbers in the cited table refer just to the number of count imputations or include other imputations. But see Beresford (A.R. C01188-90). Moreover, the Bureau's records do not reveal the exact numbers in this category but only the percentage.⁶

1970 Census

42. In 1970, the Census Bureau introduced the mail-out/mail-back procedure, allowing

⁶ The accuracy or source of the numbers cited at A.R. C01189 cannot be confirmed.

for "self-enumeration" through completion of a written questionnaire.⁷ Bureau of the Census, United States Dep't of Commerce, Report to Congress—The Plan for Census 2000, at 1 (Aug. 1997) (A.R. C00133) (hereafter "1997 Report to Congress"); Innovations, at 10, 41-43 (A.R. C01353, 1384-86). In 1970 and each subsequent census, the largest share of the population has been enumerated through the mail-out/mail-back procedure. Id. Along with the mail procedures, the Census Bureau developed a mailing list prior to Census Day. Id. at 10-11 (A.R. C01353-54). (Previously, enumerators had created a list of housing units as part of the interviewing process. Id.) With this address list, the Bureau was able to check and verify addresses prior to Census Day to achieve greater accuracy. Id.

43. In the processing of the data, the 1970 Census used imputation techniques similar to those used in the 1960 Census, except that the Census Bureau made far more extensive use of hot decks. Encyclopedia, at 197 (A.R. C00431); 1970 Census of Population and Housing, Procedural History (PHC(R)-1) 15-65 (1976) (A.R. C00395). It appears that the Census Bureau performed both household size and occupancy imputation in the 1970 Census. The 1970 apportionment count included about 900,000 imputed persons.⁸ R. Killion, DSSD Briefs, Information, and Topics Memorandum Series #F-1, Sampling and Statistical Methods in Past Censuses 4 (Jan. 13, 1997) (hereafter "Sampling and Statistical Methods in Past Censuses (#F-

⁷ In 1960, questionnaires were mailed out but individuals were instructed to turn them in to an enumerator who would be visiting their household.

⁸ Since imputation is not sampling, this number should not be included in the number that Dr. Wolfson attributes to the use of "sampling methods" in the 1970 Census. Wolfson Decl. ¶ 36 (stating 2.47 million were added to the apportionment count "based on . . . sampling methods").

1)") (A.R. C01463).

44. In addition, in the 1970 Census, two unusual situations arose that resulted in the use of imputation *in conjunction with* sampling to add persons to the apportionment count. See 1970 Census of Population and Housing, Evaluation and Research Program, Effect of Special Procedures to Improve Coverage in the 1970 Census (PHC(E)-6) Chaps. VI & VII (1976) (A.R. C00404-11); Sampling and Statistical Methods in Past Censuses (#F-1), at 2-4 (A.R. C01461-63). These procedures, known as the National Vacancy Check and the Post Enumeration Post Office Check ("PEPOC"), are described in more detail in the cited documents. Neither procedure has been used on a sampling basis in subsequent censuses. See Sampling and Statistical Methods in Past Censuses (#F-1), at 2 (A.R. C01461).

45. As can be seen by the description in the cited sources, Dr. Wolfson's description (§ 33) of the above procedures as "two distinct imputation programs, both of which were based on probability sampling," is incorrect. These were not "imputation programs." Sampling and imputation were conducted at two distinct steps of the procedures. Thus, in both cases, sampling techniques were *first* used to determine how many households had been missed, and *then* imputation techniques were used to determine population count and characteristic data for those households. See Sampling and Statistical Methods in Past Censuses (#F-1), at 2-4 (A.R. C01461-63).

1980 Census

46. The 1980 Census included both household size imputation and occupancy imputation (but did not include the sampling techniques used in the National Vacancy Check and PEPOC in 1970). See 1980 Census of Population and Housing: History, Part D, Ch. 6, at 6-28,

6-30 (1986) (A.R. C00415, C00418) (hereafter "1980 Procedural History"); D. Harner et al., Imputation in the 1980 Census (Nov. 4, 1982) (A.R. C00616); Bailar Aff. ¶¶ 8-9 (A.R. C00569); Nye Statement, at 12 (A.R. C00679). As in 1960, count imputation appears to have been used in instances of mechanical failure. 1980 Procedural History, at 12-4; D. Harner, Census Allocation Program Evaluation - Proposed Study Plan (July 28, 1981) (A.R. C00547-54).

47. As in prior censuses, imputation of both household size and occupancy status was based on geographic proximity. The specific procedures are described in more detail in the Bailar Affidavit, paragraphs 7-9 (A.R. C00568-69). In 1980, approximately 761,000 persons were added to the final count through imputation. Imputation of Population Counts (A.R. C00529-32).

48. Dr. Wolfson asserts that Dr. Bailar stated that, in the 1980 Census, the Bureau had established the "physical existence of [all] housing units' that were subjected to imputation in the 1980 Census." Wolfson Decl. ¶ 43. This statement is incorrect. Dr. Bailar merely stated that imputation "recognizes the physical existence of the housing units," a true statement. Bailar Aff. ¶ 16 (A.R. C00571). As discussed elsewhere in this declaration (¶¶ 15, 58-60, & 68-69), the Bureau has some evidence of the physical existence or validity of all addresses subject to status imputation.

49. In my Interrogatory Responses submitted in this case, I stated that "available records do not indicate for certain" whether status imputation occurred in the 1960, 1970, and 1980 Censuses. For the reasons explained in the previous paragraph, this statement is consistent with Dr. Bailar's affidavit. In any event, for the reasons explained below, the recorded use of

status imputation began with the 1990 Census, which was the first census to use a centralized address list. Based on additional information that has come to my attention since my completion of the Interrogatory Responses, which has revealed no evidence of status imputation prior to 1990, I now believe that status imputation was not performed prior to 1990. See discussion infra ¶¶ 54-55. I will be submitting shortly revised Interrogatory Responses that reflect this new information.

50. The State of Indiana sued the Census Bureau after the 1980 Census because the use of count imputation shifted a seat from Indiana to Florida.⁹ Orr v. Baldrige, No. IP-81-604-C, slip op. (D. Ind. July 1, 1985) (A.R. C00652-65). The parties stipulated that imputation was not sampling, and the Court upheld the use of imputation, holding that "the Bureau's use of hot deck imputation was an entirely reasonable means of dealing with the problem of incomplete data" Id. at 12 (A.R. C00665).

51. As part of its normal post-census evaluation process, the Census Bureau has generally researched its use of both characteristic and count imputation to determine whether imputation improves accuracy, and how the Bureau might improve imputation techniques in the future. See, e.g., C. Jones et al., Closeout Definitions, Procedures and Recommendations for Their Use in 1970 (Feb. 23, 1968) (A.R. C01204-06, 1209-11); W. Edwards Deming, The Elimination of Unknown Ages in the 1940 Population Census (Jan. 1942) (A.R. C00432-65); C.

⁹ Similarly, count imputation affected apportionment in the 2000 Census. The Bureau no longer has data available to determine whether count imputation affected apportionment in the 1960 or the 1970 Censuses.

Wolters, Intraclass Correlation Study – Analysis of 1976 RAV Data (July 31, 1978); Imputation of Population Counts (A.R. C00529-30); J. Curry, Background Information on Indiana Suit (July 28, 1981) (A.R. C00555-57); D. Harner, Census Allocation Program Evaluation - Proposed Study Plan (July 28, 1981) (A.R. C00547-54); D. Harner, Matching the Unclassified Check Results to the Census Detail File (Nov. 16, 1981) (A.R. C00561-65); D. Harner, Proposed Matching of Unclassified Check Results (Feb. 23, 1982) (A.R. C00592-97); D. Harner, Status Report of Unclassified Match (Feb. 24, 1982) (A.R. C00598-600); D. Harner, Results of Imputation Run Length Tabulation (Sept. 17, 1982) (A.R. C00601-06); D. Harner, Results of Imputation Run Length Tabulation for Special Study EDs (Sept. 20, 1982) (A.R. C00607-15); D. Harner et al., Imputation in the 1980 Census (Nov. 4, 1982) (A.R. C00616-27); K. Thomas et al., Intraclass Correlations Using a Sample of 1980 Census Data (Jan. 31, 1984) (A.R. C00634-50); J. Thompson, 1990 Census Evaluation Overview – Imputation Bias Study (Feb. 6, 1992); D. Griffin, Imputation in the 1990 Census Coverage Implications (Aug. 31, 1992) (A.R. C00734-45); J. Clark, Definitions of Imputation, Hot-Deck and Allocation (Sept. 3, 1992) (A.R. C00731); E. Schindler, Update on Study Plan for Conversion Versus Imputation Empirical Study (Sept. 16, 1992) (A.R. C00721-28); J. Clark, Definition for 1990 Imputation Evaluation (Nov. 5, 1992) (A.R. C00729-30); J. Thompson, Evaluation Plan for Imputation Bias Study (Feb. 3, 1993) (A.R. C01411-15); K. West, Imputation Bias Study: Final Report (Aug. 5, 1994) (A.R. C00748-71); J. Farber et al., A Comparison of Alternative Estimation Methodologies for Census 2000 (A.R. C01642-46); T. Williams, Comparison between the Model-Based and the Hot-Deck Methods of Imputing the Age of a Person using Data from the Sacramento 1998 Census Test Site (Mar. 4,

1999) (A.R. C01505-12). These studies collectively confirmed the validity of the count imputation methodology, the homogeneity assumption, the efficacy of the hot-deck technique used, and the overall improvement to accuracy.

1990 Census

52. Count imputation in the 1990 Census differed from 1980 in three major ways. First, count imputation was reduced to the point that it was almost non-existent. Only approximately 53,600 individuals were imputed in 1990, as opposed to about 761,000 in the 1980 Census and about 900,000 individuals in the 1970 Census. See Sampling and Statistical Methods in Past Censuses (#F-1), at 4 (A.R. C01463).

53. One reason for the very low rate of count imputation in 1990 relative to 1980 lies in the second major difference, that the Census Bureau instituted better questionnaire control procedures and more clerical edit procedures. See Statement of Stephen E. Fienberg, Maurice Falk Professor of Statistics and Social Sciences, Dean of the College of Humanities and Social Sciences, Carnegie Mellon University, before the Subcomm. on Census and Population, House Comm. on Post Office and Civil Serv., at 6 (Feb. 21, 1991) (A.R. C00688); D. Griffin, Imputation in the 1990 Census Coverage Implications 6, 8 (Aug. 31, 1992) (A.R. C00739-41).¹⁰ These

¹⁰ The "CCF" referenced by Griffin was the Census Control File, an address file maintained in the field that could be compared to the Address Control File ("ACF"), the predecessor to the MAF, to detect address problems at an earlier stage than in 1980. In addition, depending on the office where the enumerator follow-up questionnaires originated, returns received either an automated or a clerical edit, followed by a telephone call if discrepant data were found. Encyclopedia, at 197 (A.R. C00431). This clerical edit path gave clerks more human review of questionnaire forms and more opportunity to make judgments in resolving questionnaire discrepancies using prepared job aids. As a result, it appears that more discrepancies were

procedures resulted in a large drop in the number of housing units eligible for imputation. Id.

54. The third major difference relating to count imputation in 1990 was the introduction of status imputation (see ¶ 14 supra). Prior count imputation procedures had included only household size imputation and occupancy imputation. The 1990 imputation procedures continued the prior practice of using household size imputation and occupancy imputation but added status imputation. See 1990 Imputation Specifics (BB-11), at 3 (A.R. C00778). Dr. Wolfson is thus incorrect when she suggests (¶ 42) that status imputation was used for the first time in Census 2000.

55. The addition of status imputation was required because of the introduction of a centralized, computerized address list, called the Address Control File ("ACF"). Address registers for prior censuses had been maintained by hand in local census offices for each separate enumeration district, and no centralized address list was maintained. The 1990 ACF, in contrast, was a centralized, automated list of all housing units that was prepared in advance of the census. The existence of the ACF required reconciliation of each address having incomplete or conflicting information. If one source (the CCF, for example) indicated that a housing unit was a valid, nonduplicated unit, while the ACF indicated the unit should be deleted, this discrepancy had to be resolved. DSSD REX 1990 Memorandum Series #BB-7, Imputation in the 1990 Census - Coverage Implications, at 6 (Sept. 4, 1992) (A.R. C00739) (hereafter "BB-7"). The 1990 imputation procedures, therefore, had to allow for the imputation of a status of "delete," rather

resolved, and fewer returns were referred for imputation.

than only “occupied” or “vacant,” to resolve these discrepancies.

56. The detailed specifications for the imputation process in 1990 are set out in 1990 Imputation Specifics (BB-11).¹¹ In her declaration (§ 43), Dr. Wolfson points out that BB-7 (A.R. C00733-45) and another memorandum (C. Jones, Imputation of Counts in the 1990 Census (Jan. 8, 1991) (A.R. C00682)) do not mention status imputation. I have confirmed with the author of 1990 Imputation Specifics (BB-11), James B. Treat, and with individuals involved in the preparation of the other memoranda that 1990 Imputation Specifics (BB-11) accurately documents the 1990 imputation specifications and that the prior memoranda failed to discuss status imputation because the issue had not yet been thoroughly explored and because the number was trivial.

57. The Census Bureau’s review of count imputation after the 1990 Census differed from the more comprehensive review conducted after the 1980 Census because the Bureau concluded that count imputation had such a minimal effect on census accuracy in 1990 that its further study was probably not warranted. See BB-7. Additionally, further study of imputation procedures in past censuses became increasingly irrelevant because, from the early 1990s, the Bureau intended to use sampling for nonresponse follow-up in Census 2000, a plan that would have obviated the need to conduct a separate count imputation process. In paragraphs 61-75

¹¹ BB-11 was prepared by James Treat, a mathematical statistician, as part of the post-census evaluation program following the 1990 Census. Mr. Treat was instructed to form a group of knowledgeable Census Bureau employees, the Imputation Algorithm Working Group, to review and better document the 1990 imputation specifications. The Working Group reviewed the details of the 1990 imputation program and produced BB-11.

infra, I discuss the planning and design issues involved in Census 2000 in more detail. First, however, I will describe the imputation procedures and results of Census 2000.

Census 2000

58. With several exceptions noted below, the imputation procedures in Census 2000 were similar to earlier procedures. Thus, household size imputation and occupancy imputation were used, just as they had been used in 1960, 1970, 1980, and 1990. 2000 Imputation Specifications (Q-34), at 2 (A.R. C00859). In addition, as in 1990, status imputation allowed for the imputation of the existence or nonexistence of a separate, non-duplicated housing unit in instances where records were not conclusive about the unit.¹² Id. Again, all count imputation was performed through a variant of the hot-deck method. Id.

59. The Census 2000 imputation procedures introduced several modifications. The details of the procedures used are set forth in 2000 Imputation Specifications (Q-34) (A.R. C00858-86) and are summarized below:

- The Census Bureau selected the donor housing units only from the universe of housing units that had their questionnaire completed by an enumerator, rather from the universe of all housing units. A.R. C00861. For mailback areas this meant that donors were restricted to those households requiring field follow-up. Also, if

¹² The cited memorandum describes this category as containing "units for which we know nothing" and "units identified with unknown status of occupied, vacant, or delete." 2000 Imputation Specifications (Q-34), at 1-2 (A.R. C00858-59). This memorandum was written primarily for Bureau insiders, and these descriptions are intended to be shorthand descriptions of addresses that are recorded in the DMAF but for which Bureau records reflect no final status of occupied, vacant, or delete.

possible, the donor and the donee were both from a single unit structure or a multi-unit structure. A.R. C00862. The Census Bureau made this change because housing units in the follow-up universe are generally more similar to each other than to housing units that mail back their returns and households in single unit structures are more similar to each other than to those in multi-unit structures.

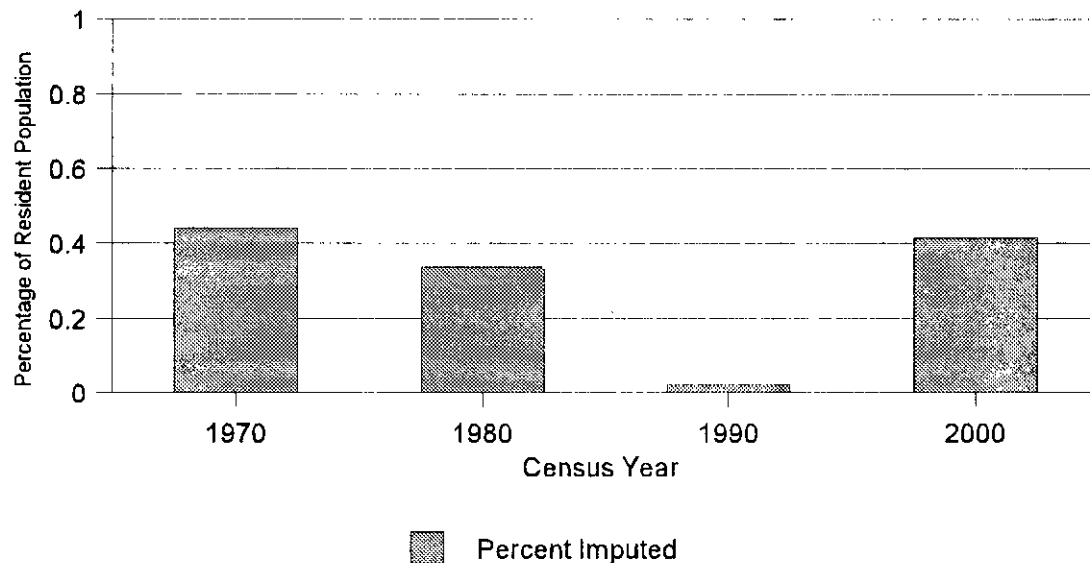
- A more sophisticated procedure was adopted to locate the donor unit. Rather than imputing from the nearest unit that was processed prior to the donee unit, the computer searched for the nearest acceptable (as defined under the first bullet above) donor unit from all processed units in the tract, rather than only those already processed. A.R. C00870.
- In 1990, the population count was capped at seven persons, the number of persons allowed on census mail return forms; that is, no housing units with greater than seven residents were used as donor housing units. This cap was dropped in Census 2000 as no rationale supported the imposition of an arbitrary cap on the size of the donor unit.
- The clerical editing process was greatly reduced in Census 2000, replaced by increased computer editing that was less subjective, more uniformly applied, and more uniform across areas and groups.

60. A total of approximately 1.2 million persons, or 0.4 percent of the population, were added to the apportionment count in 2000 through count imputation. DSSD 2000 Procedures and Operations Memorandum Series B-17, Census 2000: Missing Housing Unit

Status and Population Data (Feb. 28, 2001) (A.R. C01010-28). Although the number and percentage of count imputations were higher in Census 2000 than in 1990, the percentage was in line with the percentages in earlier censuses, as shown in the chart below.

Percent of Population Added Through Count Imputation, 1970-2000¹³

For the United States, individuals added to the apportionment totals as a result of count imputation, expressed as a percent of the total census resident population.



¹³ Data on count imputation in the 1960 Census are not included because they are not available. However, I have no reason to believe that the 1960 percentages would be notably different than those of the 1970, 1980 or 2000 Censuses. Additionally, please note that the figures from the 1970 Census include only individuals added through count imputation, not the 1.5 million individuals added through PEPOC and the National Vacancy Check, discussed in paragraph 44.

Planning for Census 2000

61. After the 1990 Census, the Census Bureau, under the direction of Congress and with input from the National Academy of Sciences, various advisory groups, and stakeholders, began to review every aspect of the cCnsus to determine how to proceed with the 2000 Census. 1997 Report to Congress, at ix-x, 7-9 (A.R. C00131-32, 139-41). Many alternatives were considered, including a sample census, a census based on administrative records, and other options. The Census Bureau's initial plan involved the use of sampling to improve accuracy by supplementing, but not replacing, initial comprehensive attempts to locate and enumerate every individual and housing unit. The plan that was adopted involved the use of sampling in three programs – nonresponse follow-up, the postal vacancy ("UAA") check, and integrated coverage measurement. See Census 2000 Operational Plan (July 1997) (A.R. C0001-125).

62. In January 1999, the Supreme Court ruled that the Census Act precluded the use of sampling to arrive at the apportionment count. Department of Commerce v. United States House of Representatives, 525 U.S. 316 (1999). As a result, the Bureau was forced to eliminate all three sampling programs from the 2000 Census. Since these methods had formed the backbone of the original census plan, the Bureau was forced to design an entirely new plan, based on its contingency plan.

63. The revised plan included, as in 1960-90, count imputation as a method to deal with missing or discrepant data. Census 2000 Operational Plan (Jan. 1999), at II-6, XI-4-5 (A.R. C001752, 1818-19) (mentioning imputation); Census 2000 Update (Feb. 1999), at 9 (same) (A.R. C01731); Census 2000 Operational Plan (Dec. 2000), II-5, XI-4-5 (A.R. C00208, 290-91)

(same). The Bureau considered count imputation to be part of its "traditional" census-taking method, as it had relied on count imputation for decades.¹⁴ Moreover, the Bureau's consistent position, articulated clearly by Dr. Bailer in the 1980 litigation, is that imputation is not sampling, and the Bureau therefore did not consider that the Supreme Court's decision might be interpreted to bar the use of imputation.

64. Imputation was not the only statistical methodology incorporated in the final plan for Census 2000. Statistical methods underlie every aspect of a modern nonsampling census, from address building to unduplication to quality assurance. Some of the statistical processes incorporated in the Census 2000 include many of the quality assurance programs, the primary selection algorithm, disclosure avoidance, the duplicate housing unit operation, and characteristic imputation. These programs are described in more detail in 2000 Quality Prototype (B-3) (A.R. C00887-937), and DSSD Census 2000 Procedures and Operations Memorandum Series B-1*, Data and Analysis to Inform the ESCAP Report 20-21 (Mar. 1, 2001) (A.R. C01088-89). For example, Census's quality assurance programs use statistical methodologies to detect possible enumerator problems and target these problem cases for re-check. 2000 Quality Prototype (B-3) (A.R. C00887-937). As another example, the multiple response resolution primary selection algorithm ("PSA") identifies multiple responses for the same census identification number and

¹⁴ Dr. Wolfson's division (§ 16) of the 2000 census-taking process into two methods, the first being "counting persons by traditional methods of enumeration" and the second being imputation, (as well as her continual use of "traditional methods" to exclude imputation) is thus incorrect and, moreover, does not take into account other statistical methods used by the Bureau.

uses a set of pre-defined criteria to determine which housing unit and person data to include in the apportionment numbers.

65. Prior to commencement of the imputation operation, the Census Bureau finalized its specifications for using imputation, DSSD Census 2000 Procedure and Operations Memorandum Series #Q-2, Census 2000 Overview of Unclassified Estimation (Aug. 8, 1999) (A.R. C00816-18); 2000 Imputation Specifications (Q-34) (A.R. C00858-86), and made public its entire census plan, including its intention to use imputation and other statistical methodologies. All details of the plan for Census 2000 were presented to the Congressional committees charged with oversight of Census 2000, the General Accounting Office, the Census Monitoring Board, the Inspector General of the Department of Commerce, and numerous advisory committees. See, e.g., Census 2000 Operational Plan (Jan. 1999), at II-6, XI-4-5 (A.R. C001752, 1818-19) (mentioning imputation); Census 2000 Update (Feb. 1999), at 9 (same) (A.R. C01731); Census 2000 Operational Plan (Dec. 2000), II-5, XI-4-5 (A.R. C00208, 290-91) (same). Count imputation was used in the 1998 Dress Rehearsal at the site that tested the traditional census plan (the South Carolina site), and the imputation numbers from this test were disclosed in professional articles. See R. Singh et al., Census 2000 Dress Rehearsal Methodology and Initial Results (A.R. C01657-62). This article was dated August 1999 and supplied to the Census Monitoring Board in November, 1999. Many oversight groups intensely scrutinized the planning and conduct of Census 2000, *yet no one commented on, questioned, or opposed the continued use of count imputation.*

Census 2000 Operations

66. In 2000, the Bureau went to great lengths to attempt to obtain responses from every household, focusing on outreach and instituting new procedures and programs to improve the mail response rate and to provide other convenient ways to respond to the census. Census 2000 Operational Plan (Dec. 2000), at I-1--3, V-7 (A.R. C00198-200, 233). The Census Bureau designed and implemented an enhanced marketing and partnership program to provide an integrated communications effort to increase both awareness of the decennial census and public cooperation. Id. § IV (A.R. C00220-24). The marketing program was designed around the first-ever paid advertising campaign, including a national media campaign aimed at increasing mail response, targeted advertising directed at raising mail response among historically undercounted populations, and special advertising messages and campaigns targeted to hard-to-enumerate populations. In the partnership program, the Census Bureau worked nationwide with state and local partners to encourage all individuals to respond to the census.

67. Perhaps just as important, the Census Bureau worked with private sector experts to redesign its mail response strategy, preparing forms that were easier to complete and increasing the number of mail contacts with respondents. Census 2000 Operational Plan (Dec. 2000), at I-2--3, V-7 (A.R. C00199-200, 233). The content of the short-form questionnaire was redesigned, asking only seven simple questions. Id. at V-6 (A.R. C00232). Also, the Census Bureau instituted, for the first time, an extensive multiple mailing strategy using first-class mail, first sending an advance letter to every mail-out address, alerting the residents that the census form will be sent soon, followed by the questionnaire itself, and then a postcard that served as a "thank

you" for those who responded and a reminder for those who did not. Id. at V-7 (A.R. C00233).

68. Census accuracy is tied closely to an accurate address file, now called the Decennial Master Address File ("DMAF"). In preparation for Census 2000, the Census Bureau completed an intensive Census 2000 Address List Reengineering effort, making changes in the DMAF design that were expected to increase its accuracy and completeness. Census 2000 Operational Plan (Dec. 2000), § VI (A.R. C00244-56). Among other things, the Bureau combined the 1990 address list with data provided by the Postal Service's automated Delivery Sequence File to create the DMAF. Addresses were also added to the DMAF during a pre-census canvass by enumerators. Finally, a new program, the local update of census addresses program ("LUCA"), conducted pursuant to 13 U.S.C. § 16, gave state, local, and tribal governments an opportunity to provide input regarding the completeness and accuracy of the DMAF. As the census date approached, the Bureau continued to revised and update the DMAF. See DSSD Census 2000 Procedures and Operations Memorandum #D-6, Specification for Updating the Decennial Master Address File in April, 2000 (Mar. 27, 2000) (A.R. C01545-54); DSSD Census 2000 Procedures and Operations Memorandum Series #D-11, Specification for Reinstating Addresses Flagged as Deletes on the Hundred Percent Census Unedited File (HCUF) (Nov. 7, 2000) (A.R. C01579-84) (hereafter "Reinstatement Specification"); see generally The Census Bureau's Master Address File (MAF): Census 2000 Address List Basics (Mar. 1999) (A.R. C01490-504).

69. Dr. Wolfson states in paragraph 27 of her Declaration that the "Bureau's 'double-delete' policy required two independent verifications of vacancy or non-existence for a household to be given a final classification status of 'vacant' or 'delete' (i.e., non-existent). . . . Therefore, in many instances the Bureau presumably 'counted' persons living in units whose occupancy or existence was not only unverified, but also doubtful." This is incorrect. The "double-delete" rule was only one criterion used to assign a final status of "delete." There are several other criteria that led to the assignment of a "delete" status when the Bureau had a response that indicated an address was not valid or a duplicate. For example, if the Bureau had not received a mail return from a unit and the enumerator response indicated that the unit was nonexistent, the housing unit was given a final status of delete. In all instances, housing units were given a final status of vacant or delete whenever the questionnaire indicated in a consistent fashion that the housing unit was vacant or nonexistent. DSSD Census 2000 Procedures and Operations Memorandum Series #D-14, Specifications for Assigning the Housing Unit Status and Population Count on the Hundred-Percent Unedited File Prior to the Imputation for Unclassified Units (Jan. 19, 2001) (A.R. C00988-1009); DSSD Census 2000 Procedures and Operations Memorandum Series #D-13, Specification of the Kill Universe on the Decennial Master Address File for Census 2000 (Dec. 21, 2000) (A.R. C01585-1601); Reinstatement Specification (A.R. C01579-84).¹⁵

¹⁵ These procedures were just as stringent, if not more so, in 1990. See General Government Division, United States Government Accounting Office, GAO /GCD-92-94, Decennial Census: 1990 Results Show Need for Fundamental Reform 25 (June 1992) (Ex. F hereto) ("To help ensure that the [address] list would be as complete as possible, once an address was placed on the list, it generally was not removed unless identified for deletion in at least three address list development or census follow-up operations.").

70. The Census Bureau used four basic data collection methods in Census 2000. Census 2000 Operational Plan (Dec. 2000), at IX-1, IX-8--9 (A.R. C00267, 274-75). In the two most common procedures, respondents were asked to mail back their questionnaires: in mail-out/mail-back, the Bureau mailed questionnaires to all housing units on the DMAF for the residents to mail back and, in update/leave, Census Bureau workers delivered questionnaires to addresses on the DMAF for residents to mail back and updated the address list. In the two other procedures, households were interviewed directly: in list/enumerate, Census Bureau enumerators created the address list while canvassing their assignment areas and conducted interviews with respondents, and in update/enumerate, enumerators began with an address list, updated it, and conducted the interviews. These last two procedures were used in remote, distant, and inaccessible areas.

71. In addition, the Census Bureau developed unduplication technology that allowed it, for the first time, to permit multiple methods of responding to the Census. Census 2000 Operational Plan (Dec. 2000), at I-3, IX-14 (A.R. C00200, 280). These methods for responding to the Census were not possible in 1990 because the Bureau's matching and unduplication technologies were not adequate to permit unduplication of responses (to make sure individuals were not counted twice – by one spouse sending in the questionnaire and the other calling the Census Bureau, for example). Thus, in Census 2000, some individuals were able to respond to the Census through the Internet. The Census Bureau established Telephone Questionnaire Assistance centers that respondents could call to obtain help in filling out their forms and to complete their response over the telephone if they wished. And the Be Counted program

provided an additional means for people to be included in the Census by allowing them to fill out a blank form made available in various public locations. Id. at IX-2--3 (A.R. C00268-69).

72. Finally, special enumeration procedures were followed for remote parts of Alaska, for locations containing a concentration of persons with a transient lifestyle (e.g., trailer parks, marinas, and campgrounds), for group quarters (e.g., prisons and long-term care facilities), and for people with no usual residence. Census 2000 Operational Plan (Dec. 2000), at IX-4--7 (A.R. C00270-73).

73. After allowing a reasonable amount of time for respondents to mail back their questionnaires, the Census Bureau conducted a nonresponse follow-up ("NRFU") operation, a field follow-up of housing units that had not returned their questionnaires by mail. During NRFU, census enumerators made up to six attempts to contact housing units that appeared occupied to secure an interview (those that appeared vacant were recorded as such and followup efforts were ceased). United States Census 2000: Final Attempt Procedures for FOSs (Nonresponse Followup) (Apr. 2000) (A.R. C00819-41). If the enumerator could not obtain an interview, he or she attempted to interview a proxy respondent, that is, a neighbor, rental agent, building manager, or other knowledgeable individual. Census 2000 Operational Plan (Dec. 2000), at IX-12 (A.R. C00278); Feasibility Statement, at 8 (A.R. C01910).

74. A number of additional operations were implemented to ensure as complete coverage as possible in the initial enumeration. Computer edits were performed on mail-return questionnaires to identify returns that contained missing persons, missing person information, or large housing units (more than six persons). Interviewers then conducted telephone interviews

with those housing units during the Coverage Edit Follow Up ("CEFU") operation to obtain accurate data about the persons residing there. Another operation, Coverage Improvement Follow Up ("CIFU"), was conducted after NRFU. This operation required an interviewer recheck of housing units classified as vacant or nonexistent during NRFU to ensure that no units were misclassified as well as to check on new construction, lost or blank forms, and other coverage problems. This operation enumerated 2.3 million occupied housing units. Finally, after the data collection efforts were completed, the data were processed through a number of computer operations to unduplicate multiple responses for the same housing unit and to edit inconsistent or missing responses. Census 2000 Operational Plan (Dec. 2000), at IX-10-11, 13-14 (A.R. C00276-77, 279-80).

75. Finally, major operations of the Census 2000 plan were subjected to quality assurance activities designed to detect and correct errors before they affected accuracy or data quality. 2000 Quality Prototype (B-3) (A.R. C00887-937).

76. In spite of these intense efforts to locate and enumerate every individual, inevitably housing units existed for which the Census Bureau did not have sufficient data, requiring the imputation procedures described above.

77. Following the 2000 Census, the Census Bureau, as it does after all censuses, is still conducting a variety of studies and analyses to assess the effectiveness and accuracy of the various procedures used in the census. One such effort is addressed to the use of count imputation. Specifically, the Census Bureau assembled an interdivisional team at the beginning of January 2001 to analyze, assess, and document the imputation results from Census 2000.

78. This team has not yet completed its work. However, I have reviewed preliminary results for the purpose of this declaration. These preliminary results confirm that most of the count imputations performed in Census 2000 are attributable to housing units that have been determined to exist, but whose data were not included in the totals through a variety of processing errors. These cases appear to have been appropriately included in the census. If they had not been included through count imputation, these cases would represent individuals or housing units that attempted to be included in the Census and who were left out simply through processing errors.

I declare under penalty of perjury that the foregoing is true and correct.

July 10, 2001
DATE

Howard Hogan
HOWARD HOGAN, Ph.D.
Chief, Decennial Statistical Studies Division
Bureau of the Census
United States Department of Commerce

STUART E. SCHIFFER
Acting Assistant Attorney General
SANDRA M. SCHRAIBMAN
CAROL FEDERIGHI
AMY M. ALLEN
Attorneys, U.S. Department of Justice
Civil Division, Federal Programs Branch
P.O. Box 883, 901 E Street, N.W.
Washington, D.C. 20044
Telephone: (202) 514-4523
Facsimile: (202) 616-8202
Attorneys for Defendants

**IN THE UNITED STATES DISTRICT COURT
DISTRICT OF UTAH, CENTRAL DIVISION**

STATE OF UTAH, et al.,

Plaintiffs,

v.

**DONALD L. EVANS, United States Secretary
of Commerce, et al.**

Defendants.

)
)
) **Civil No. 2:01-CV-000292G**
)
)
)

) **EXHIBITS TO DECLARATION OF**
) **HOWARD HOGAN**
)
)
)

INDEX

| <u>Exh.</u> | <u>Document</u> |
|-------------|--|
| A | Curriculum Vitae |
| B | General Government Division, United States General Accounting Office, GAO/GCD-98-103, <u>Decennial Census: Overview of Historical Census Issues</u> (May 1998) (excerpts) |
| C | Innis G. Sande, <u>Hot-Deck Imputation Procedures, in 3 Incomplete Data in Sample Surveys</u> 339 (W.G. Madow & I. Olkin eds., 1983) |
| D | Frederick J. Scheuren, <u>Discussion, in 3 Incomplete Data in Sample Surveys</u> 353, 357 (W.G. Madow & I. Olkin eds., 1983) |
| E | Barry L. Ford, <u>An Overview of Hot-Deck Procedures, in 2 Incomplete Data in Sample Surveys</u> 185 (W.G. Madow & I. Olkin eds., 1983) (excerpts) |
| F | General Government Division, United States Government Accounting Office, GAO /GCD-92-94, <u>Decennial Census: 1990 Results Show Need for Fundamental Reform</u> (June 1992) (excerpts) |

Exhibits/
Attachments
to this document
have **not** been
scanned.

Please see the
case file.